Amendments to the Specification

Please replace the paragraph starting at page 2, line 14, with the following amended paragraph:

- Ectopic expression of ~~HLA~~ human leukocyte antigens (HLA) class II antigens: class II antigens have a restricted tissue distribution. The tissues affected in autoimmune diseases may express class II antigens inappropriately.

Please replace the paragraph starting at page 4, line 17, with the following amended paragraph:

As used herein, the term "co-segregation" relates to any association of the mutated form of the polypeptide with APECED. APGD1 is a protein with a predicted length of 545 amino acids, a theoretical molecular weight of ~~57,7~~ 57.7 kD and a calculated pI of ~~7,53~~ 7.53. Statistical analysis of the protein sequence of Fig. 2A (Brendel, V., et al., *Proc. Natl. Acad. Sci. USA,* **89**, 2002-2006 (1992)) indicates a high content of proline (11.7%) but no apparent clusters of charged amino acids or periodicity patterns. The secondary structural content of APGD1 was predicted to consist mostly of coils, with only a weak probability for the occurrence of structural α-helixes or β-sheets. A putative bi-partite nuclear targeting signal (Dingwall, C. & Laskey R. A.,*TIBS,* **16**, 478-481 (1991)) was found between amino acids 113 to 133 (Figure 2A). The predicted protein harbors two cysteine-rich regions of 42 amino acids, each specifying a Cys4-His-Cys3 double-paired finger motif similar to the PHD finger type (Aasland, R., et al., *TIBS,* **20**, 56-59 (1995)) (Figure 2A). Spacing of essential residues is conserved in the two motifs found in APGD1: $C_{299,434}$-XX-$C_{302,437}$-X(8)-$C_{311,446}$-XX-$C_{314,449}$-X(4)-$H_{319,454}$-XX-$C_{322,457}$-X(14)-$C_{337,471}$-XX-$C_{340,474}$ (where X is any amino acid and numbers in parenthesis represent the length of the intervening peptide sequence). This structural motif has been reported for a number of nuclear proteins involved in the mediation or regulation of transcription, such as TIF1 (Transcription Intermediary Factor 1) (Douarin, Le, B., et al., *EMBO J.*, **14**, 2020-2033 (1995)) and KRIP-1 (KRAB-A Interacting Protein) (Kim, S-S., et al., *Proc. Natl. Acad. Sci, USA*, **13**, 15299-14304 (1996)). Sequence homology of APGD1 with other proteins in the databases was strictly limited to this

Cys4-His-Cys3 motif. Although the spacing of residues is conserved in each case, the sequence is most closely homologous to the Mi-2 autoantigen (Ge, Q., et al., *J. Clin. Invest.*, **96**, 1730-1737 (1995)) and the TIF1 proteins (Thenot, S., et al., *J. Biol. Chem.*, **272**, 12062-12068 (1997)). Mi-2 is the major nuclear antigen detected in the sera of autoimmune dermatomyositis patients (Ge, Q., et al., *J. Clin. Invest.*, **96**, 1730-1737 (1995)) and TIF1 is involved in the transcriptional control of the estrogen receptor (Thenot, S., et al., *J. Biol. Chem.*, **272**, 12062-12068 (1997)).

Please replace the paragraph starting at page 4, line 20, with the following amended paragraph:

As used herein, the term "co-segregation" relates to any association of the mutated form of the polypeptide with APECED. APGD1 is a protein with a predicted length of 545 amino acids, a theoretical molecular weight of 57,7 kD and a calculated pI of 7,53. Statistical analysis of the protein sequence of Fig. 2A (Brendel, V., et al., *Proc. Natl. Acad. Sci. USA,* **89**, 2002-2006 (1992)) indicates a high content of proline (11.7%) but no apparent clusters of charged amino acids or periodicity patterns. The secondary structural content of APGD1 was predicted to consist mostly of coils, with only a weak probability for the occurrence of structural α-helixes or β-sheets. A putative bi-partite nuclear targeting signal (Dingwall, C. & Laskey R. A.,*TIBS,* **16**, 478-481 (1991)) was found between amino acids 113 to 133 (Figure 2A). The predicted protein harbors two cysteine-rich regions of 42 amino acids, each specifying a Cys4-His-Cys3 (SEQ ID NO.:29) double-paired finger motif similar to the PHD finger type (Aasland, R., et al., *TIBS,* **20**, 56-59 (1995)) (Figure 2A). Spacing of essential residues is conserved in the two motifs found in APGD1: $C_{299,434}$-XX-$C_{302,437}$-X(8)-$C_{311,446}$-XX-$C_{314,449}$-X(4)-$H_{319,454}$-XX-$C_{322,457}$-X(14)-$C_{337,471}$-XX-$C_{340,474}$ (SEQ ID NO:30) (where X is any amino acid and numbers in parenthesis represent the length of the intervening peptide sequence). This structural motif has been reported for a number of nuclear proteins involved in the mediation or regulation of transcription, such as TIF1 (Transcription Intermediary Factor 1) (Douarin, Le, B., et al., *EMBO J.*, **14**, 2020-2033 (1995)) and KRIP-1 (KRAB-A Interacting Protein) (Kim, S-S., et al., *Proc. Natl. Acad. Sci, USA*, **13**, 15299-14304 (1996)). Sequence homology of APGD1 with other proteins in the databases was

strictly limited to this Cys4-His-Cys3 (SEQ ID NO:29) motif. Although the spacing of residues is conserved in each case, the sequence is most closely homologous to the Mi-2 autoantigen (Ge, Q., et al., *J. Clin. Invest.*, **96**, 1730-1737 (1995)) and the TIF1 proteins (Thenot, S., et al., *J. Biol. Chem.*, **272**, 12062-12068 (1997)). Mi-2 is the major nuclear antigen detected in the sera of autoimmune dermatomyositis patients (Ge, Q., et al., *J. Clin. Invest.*, **96**, 1730-1737 (1995)) and TIF1 is involved in the transcriptional control of the estrogen receptor (Thenot, S., et al., *J. Biol. Chem.*, **272**, 12062-12068 (1997)).

Please replace the paragraph starting at page 5, line 23, with the following amended paragraph:

As stated hereinabove, the invention also relates to nucleic acid molecules which hybridize to the above described nucleic acid molecules and differ at one or more positions in comparison to these as long as they encode a (poly)peptide having the above described characteristics. In connection with the present invention, the term "hybridizing" is understood as referring to conventional hybridization conditions, preferably such as hybridization in 50% formamide, 6x SSC, 0.1% ~~SDS~~ sodium dodecyl sulfate (SDS), and 100µg/ml ~~ssDNA~~ single stranded DNA (ssDNA), in which temperatures for hybridization are above 37°C and temperatures for washing in 0.1x ~~SSC~~ sodium chloride sodium citrate (SSC), 0.1% SDS are above 55°C. Most preferably, the term "hybridizing" refers to stringent hybridization conditions, for example such as described in Sambrook, et al. (Molecular cloning; A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY (1989)) or Higgins & Hames (Nucleic acid hybridization, A practical approach, IRL Press, Oxford (1985)). Said nucleic acid molecules comprise those which differ, for example, by deletion(s), insertion(s), alteration(s) or any other modification known in the art in comparison to the above described nucleic acid molecules. Methods for introducing such modifications in the nucleic acid molecules according to the invention are well-known to the person skilled in the art; see, e.g., Sambrook, et al., supra.

Please amend the paragraph starting on page 8, with the following amended paragraph:

As mentioned above, said mutation results in a loss or a gain of function of the (poly)peptide of the invention. In a preferred embodiment of the present invention, said loss of function is a loss of macromolecule binding properties. However, a loss of transactivating property in addition or instead of the loss of the macromolecule binding property is also envisaged. Other possibilities relate to the loss of a structural determinant (truncated protein) in addition to the loss of a functional determinant.

Please replace the paragraph starting on page 8, line 24, with the following amended paragraph:

For example, the experiments performed in accordance with the present invention suggest that at least some of the mutations identified so far in the AIRE gene lead to truncated forms of the (poly)peptide of the present invention lacking at least one of the PHD zinc fingers. Based on the cellular localization studies performed in accordance with the present invention (for details see Examples 10 to 12) it is, furthermore, envisaged in accordance with the present invention, but without being bound to any scientific theory, that loss of function of the mutated/truncated (poly)peptides of the invention may be associated with their abnormal nuclear distribution. Thus, it is conceivable that the truncated (poly)peptides of the invention are erroneously directed to other nuclear structures by default as consequence of missing a domain normally interacting with either a core DNA target or chromatin-associated protein. In addition, it could be shown in accordance with the present invention that AIRE interacts with structural components of the cytoplasmic compartment. More specifically, it is an envisaged that AIRE associates with vimentin since AIRE harbors a cluster of basic amino acids within the nuclear targeting signal. Moreover, the apparently variable temporal and spatial decoration of filament arrays and nuclear speckles by anti-AIRE antibodies suggests the existence of a dynamic or passive trafficking of AIRE in the cell. Thus, it is also envisaged in accordance with the present invention that AIRE is residing on vimentin fibers as part of a docking mechanism regulating nuclear translocation. The occurrence of nuclear factors

interacting with components of the cytoskeleton is not an unprecedented observation. An interesting example is the regulation of the function of *Gli* zinc finger transcription factor, vertebrate homologue of Drosophila *ci* gene (Biesecker, L.G. (1997). Strike three for GLI3 [news] [published erratum appears in Nat Genet 1998 Jan;18(1):88]. *Nature Genetics* **17**, 259-260). This transcription factor is mainly targeted to the cytoplasm where it is anchored to microtubules, whereas a truncated form of Gli processed by proteolytic cleavage of the molecule is directed to the nucleus (Aza-Blanc, P., Ramirez-Weber, F.A., Laget, M.P., Schwartz, C. & Kornberg, T.B. (1997). Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. *Cell* **89**, 1043-1053; Robbins, D.J., Nybakken, K.E., Kobayashi, R., Sisson, J.C., Bishop, J.M. & Therond, P.P. (1997). Hedgehog elicits signal transduction by means of a large complex containing the kinesin-related protein costal2. *Cell* **90**, 225-234). To date, the only described nuclear factor interacting with vimentin is a protein component of the nuclear matrix, NMP125, transiently stored along vimentin during mitosis (Marugg, R.A. (1992). Transient storage of a nuclear matrix protein along intermediate-type filaments during mitosis: a novel function of cytoplasmic intermediate filaments. *Journal of Structural Biology* **108**, 129-139). Thus, AIRE represents the first example of a zinc-finger protein co-localizing with vimentin intermediate filaments. With respect to the abnormal cytoplasmic localization, it is thus envisaged that loss of function may be associated with impaired protein-protein interactions involved in maintaining the shape and integrity of intermediate filaments. In other words, aggregates of the mutant (poly)peptides of the present invention may prevent the formation of vimentin intermediate filaments by, e.g., entrapping vimentin. On the other hand, it may also be envisaged that the above-mentioned docking/activation mechanism of the mutant (poly)peptides of the invention is impaired thereby leading to a loss of function.
Thus, the pathological consequences of at least some of the mutations found in the AIRE gene may elicit their effects at least in part by effecting the spatial organization of AIRE in the cell.

Please replace the paragraph starting on page 13, line 28, with the following amended paragraph:

Furthermore, the invention relates to a (poly)peptide encoded by a nucleic acid molecule of the invention or produced by the above described process. In this context it is also understood that the (poly)peptides according to the invention may be further modified by conventional methods known in the art. By providing the (poly)peptides according to the present invention it is also possible to determine the portions relevant for their biological activity. This may allow the construction of chimeric proteins or fusion proteins comprising an amino acid sequence derived from a (poly)peptide of the invention which is crucial for its biological activity and other functional amino acid sequences like, e.g., nuclear localization signals, transactivating domains, DNA-binding domains, hormone-binding domains, protein tags ~~GST~~ glutathione S-transferase (GST), ~~GFP~~ Green Fluorescent Protein (GFP), h-myc peptide, Flag, ~~HA~~ histocompatability antigen (HA) peptide) which may be derived from the same or from heterologous proteins. Said chimeric or fusion proteins are also comprised by the present invention.

Please replace the paragraph starting at page 17, line 1, with the following amended paragraph:

In yet another embodiment, the present invention relates to methods for testing for a carriership for APECED or for a corresponding disease state comprising testing a sample obtained from a prospective patient or from a person suspected of carrying a predisposition for a mutated form of the (poly)peptide(s) according to the invention. Such methods comprise, e.g., immuno-precipitation, immuno-blotting, ~~ELISA~~ enzyme linked immunosorbent assay (ELISA), ~~RIA~~ radioimmunoassay (RIA), indirect immuno-fluorescence experiments, and the like. Such techniques are well known in the art and are described, e.g. in Harlow and Lane, supra.

Please replace the paragraph starting on page 19, line 2, with the following amended paragraph:

A) The physical map of the APECED region showing the markers used to construct the disease haplotypes (cen - JA1, D21S1912, PFKL (CA_n), PB1, D21S171 - tel), the other genes (PFKL, green and 694N10, pink) and the ESTs (EST cluster 1: AA082879, AA085392, EST cluster 2: N67176, T84071, T86112, T79577, T79655, R23544, R44295, EST cluster 3: AA453553) located in the close vicinity of APGD1 (SEQ ID NOS.: 1-2) (blue) and the key cosmid clones Q21D11 and Q22G11 used for genomic sequencing as well as cosmid clone Q11D11 that was used as orientation marker in the fiber FISH experiment (see Figure 1C).

Please replace the paragraph starting on page 19, line 11, with the following amended paragraph:

B) The genomic structure of the APGD1 gene. The 14 true exons of the gene are compared with the gene models predicted with different gene finding programs (Uberbacher, E., et al., *Proc. Natl. Acad. Sci, USA*, **88**, 11261-11265 (1991); Thomas, A., & Skolnick, M. H., *IMA J. Math. Appl. Med. Biol.*, **11**, 149-160 (1994); Kulp, D., et al., ISMB-96, St. Louis, MO, AAAI/MIT Press. (~~http://www-hgc.lbl.gov/projects/genie.html~~ available on the worldwide web at hgc.lbl.gov/projects/genie.html) (1996)). Solid boxes indicate exons in which at least one boundary was correctly predicted, open boxes are false exons. Genomic sequence of cosmid clones Q21D1, Q22G11, EST matches, detailed gene prediction data and the intron-exon boundaries of APGD1 are available at http://chr21.rz-berlin.mpg.de/APECED.html/.

Please replace the paragraph starting on page 22, line 28, with the following amended paragraph:

Dot-matrix of sequence comparison of the human (SEQ ID NO:4) and murine SEQ ID NO.:3) AIRE gene structure (A). Arrows mark exons. Arrowhead denotes conserved region (SEQ ID NO:5) shown in detail in Figure 13B.

Please replace the paragraph starting on page 22, line 32, with the following amended paragraph:

cDNA sequence of murine AIRE (SEQ ID NO:6) gene and deduced amino acid sequence (SEQ ID NO:7).

Please replace the paragraph starting at page 23, line 9, as follows:

Amino acid sequence comparison of the human (SEQ ID NO:8) and murine (SEQ ID NO:9) AIRE protein. Consensus sequence is indicated (SEQ ID NO:10). Shaded boxes mark PHD fingers and the dolled line the SAND domain. The unclear localization signal (NLS) is underlined, and the LXXLL-motif is boxed.

Please replace the paragraph starting at page 23, line 14, with the following amended paragraph:

Differential splicing of the mouse AIRE gene. Amino acid sequence (SEQ ID NO:12) is indicated above the nucleic acid sequence (SEQ ID NO:11).

(a) Shows skipping of exon 10 (SEQ ID NO:11-12);

(b) Shows deletion of a lysine in exon 8 (SEQ ID NO:13-14);

Shows deletion of Proline, Isoleucine, Threonine, Valine in exon 6 (SEQ ID NO:15-16).

Please replace the paragraph staring at page 23, line 21, with the following amended paragraph:

Expression of human AIRE in a series of immunological tissues. ~~RT-PCR~~ Reverse transcription-polymerase chain reaction amplification was performed as described in Example 15. Lanes 1 to 8 correspond to: fetal liver, lymph node, peripheral blood leukocyte, thymus, bone marrow and spleen respectively. Lane 9 is negative control; M1 is lambda HindIII marker, M2 is 100 bp ladder marker.

Please replace the paragraph starting on page 23, line 30, with the following amended paragraph:

We have mapped APECED to chromosome 21q22.3 by linkage analysis and further refined the localisation by linkage disequilibrium to a region between the markers D21S25 and D21S171 (Aaltonen, J., et al., Nature Genet, 8, 83-87 (1994); Aaltonen et al., Genome Research 7 (1997), 820-827). This critical region was 350 kb in size and a bacterial clone contig was constructed across this region. Several techniques were used to identify candidate genes in this gene rich region. Exon trapping (Buckler, A., et al., *Proc. Natl. Acad. Sci., USA*, 88, 4005-4009, (1991)) and cDNA selection (Lovett, M., et al., *Proc. Natl. Acad. Sci, USA*, 88, 9628-9632, (1991)) methods identified a new gene, 694N10 (Accession No. Z93322), just distal to the previously known PFKL gene (Phosphofructokinase of liver type, EC 2.7.1.11) (Elson et al., Genomics, 7, 47-56 (1990)) (Figure 1A). Partial unordered genomic sequence encompassing the PFKL gene (available at the International Chromosome 21 genomic sequence repository, on the worldwide web at eri.uchsc.edu ~~http:// www-eri.uchsc. edu/chr21/ eridna.html~~) was used to generate a new polymorphic marker, PB1. This marker showed an obligatory recombination in one APECED family, thus we were able to restrict the APECED region to 145 kb between the markers D21S25 and PB1 (Figure 1A). Therefore 694N10 was excluded as causative gene for APECED.

Please replace the paragraph starting at page 24, line 25, with the following amended paragraph:

In parallel, we initiated a large scale sequencing approach from cosmid clones 21D1 and 22G11 mapping to the critical region (Figure 1A). A total of 87 kb of genomic sequence obtained from these cosmids were analysed with BlastN and BlastX algorithms (Altschul, S. F., et al., *J. Mol. Biol.*, **215**, 403-410, (1990)) against public databases. Three different EST (Expressed Sequence Tag) clusters were found in a region between D21S25 and PFKL (Figure 1A). Exon prediction was performed using the GRAIL2 program (Uberbacher, E., et al., *Proc. Natl. Acad.*

Sci, USA, **88**, 11261-11265 (1991)). A gene model was predicted directly upstream of the promotor of PFKL where no EST matches were identified (exons G1 to G7, Figure 1B). However, since the linkage disequilibrium data (Björses, P., et al., *Am. J. Hum. Genet.*, **59**, 8779-886 (1996)) suggested the APECED gene to be located in the close vicinity of PFKL further analyses were focused on this potential gene. Polymerase Chain Reaction (PCR) amplification (5'- AGA AGT GCA TCC AGG TTG GC-3' (SEQ ID NO. 17) and 5'-GGA AGA GGG GCG TCA GCA AT-3') (SEQ ID NO:18) of a 316 bp genomic fragment spanning predicted exons G5 and G6 (Figure 1B) generated a probe for screening a human adult thymus cDNA library (Clontech catalog # HL5010b). Two cDNA clones (B1-1 and D1-1) and a 3' UTR extension PCR product yielded a composite cDNA sequence of 2,245 kb (Figure 2A). The cDNA clone B1-1 was localised on the physical map by fiber FISH (Fluorescent In Situ Hybridization) (Figure 1C) (Heiskanen, M., et al., *TIG*, **10**, 379-382 (1996)). Northern blot analysis showed a major transcript of approximately 2 kb expressed in all tissues analysed, the most intensive signals were obtained from thymus, pancreas and adrenal cortex (Figure 2B). In this respect, it is surprising that no ESTs were found in the databases. The cDNA sequence exhibits an unusually high GC content of 68.8% and contains an open reading frame (ORF) of 581 amino acids followed by a STOP codon at nucleotide 1756. The likely initiator ATG codon occurs at nucleotide 121 (Figure 2A), predicting a 545 residue protein.

Please replace the paragraph starting at page 25, line 8, with the following amended paragraph:

The structure of the APGD1 gene was determined from a comparison of the cDNA sequence with the cosmid 22G11 genomic sequence using the est_genome program (developed by Richard Mott, available at the Sanger center, UK). The genomic structure consists of 14 exons spanning 11,9 kb of genomic DNA (Figure 1B). A putative promotor containing a TATA box located 35 nucleotides from the first nucleotide of exon 1 and a GC box was identified immediately upstream of the first exon of the APGD1 gene. A CpG island was also associated with the promotor region. Detailed analysis of the genomic sequence upstream of the APGD1

gene did not suggest any additional exons within 22 kb of the predicted promotor. The translation of the genomic sequence identified an in frame STOP codon 16 residues upstream of the first amino acid of the translated cDNA sequence. Analysis of the 3' end of the gene suggested that exon 14 represents the last exon since the STOP codon at position 1756 is followed by repetitive sequences. Further, exon 14 overlaps with the promoter region of the PFKL gene (Levanon, D., et al., *Biochem and Mol. Biol. Int.*, **35**, 929-936 (1995)) which is transcribed from the same DNA strand (Figure 1B and 2A). Apparent C to T silent polymorphisms were found at third codon positions in exons 5, 6, 10 and 14 (Figure 2A). The gene organisation was poorly predicted by GRAIL: only three (exons 2, 4 and 6) of the 14 exons were identified *bona fide* and 7 exons were completely missed (Figure 1B). Yet, the gene is located in a GC rich region and intron-exon boundaries follow the GT-AG rule (Mount, S.M., et al., *Nucleic Acids Research.*, **10**, 459-472 (1982)). Subsequent analysis of the genomic sequence with other gene finding software including GRAIL1a (Uberbacher, E., et al., *Proc. Natl. Acad. Sci, USA*, **88**, 11261-11265 (1991)), Xpound (Thomas, A., & Skolnick, M. H., *IMA J. Math. Appl. Med. Biol.*, **11**, 149-160 (1994)), and Genie (Kulp, D., et al., ISMB-96, St. Louis, MO, AAAI/MIT Press. (~~http://www-hgc.lbl.gov/projects/genie.html~~ available on the worldwide web at hgc.lbl.gov/projects/genie.html) (1996)) showed that Genie, based on hidden Markov model, performed best for modeling the 3' end of this gene (Figure 1B).

Please replace the paragraph starting at page 26, line 4, with the following amended paragraph:

For mutation screening in APECED patients, all 14 exons were amplified from genomic DNA using primers located in the respective flanking introns (primer sequences and the detailed protocols available on the worldwide web at ~~http://chr21.rz-berlin.mpg.de/APECED.html~~ chr21.rz-berlin.mpg.de/APECED.html). Five different mutations were identified in the coding region of APGD1 (Table 1). The mutations were monitored in a control panel of 500 unrelated Finns and 60 unrelated Europeans including 32 CEPH parents. The most common mutation was the "Finnish major mutation" found in 82% of the Finnish patients, all of which have the major

disease haplotype (No. 1.1 in Table 1) (Björses, P., et al., *Am. J. Hum. Genet.*, **59**, 8779-886 (1996)). This mutation is a C to T transition at nucleotide 889 in exon 6, changing an Arg into a STOP codon. Among the 500 Finns this mutation was detected in two heterozygotes, indicating a carrier frequency of 1: 250. The same mutation was also found in an Italian and in a German patient, who carried different haplotypes (haplotypes No. 1.2 to 1.4 in Table 1, respectively). Two mutations were found in exon 8. The first one is a duplication of four nucleotides (CCTG) normally found at position 1086 to 1089. The other mutation in this exon is a 13 bp deletion (nucleotides 1085 to 1097) observed in four non-Finnish patients (two British, a Dutch and a German) carrying the same haplotype (No. 2.1 in Table 1). Two other mutations which involve insertion or deletion of a single nucleotide were found in exon 10. The insertion of an A at position 1284 was found in two compound heterozygote Finnish patients having the Finnish major mutation in the other allele. Deletion of a C was found at position 1313 in a French patient homozygous for the disease haplotype (No. 5.1 in Table 1). Mutations and the associated haplotypes are summarized in Figure 3 and Table 1. Northern blot analysis performed on lymphoblast mRNA from patients whose cell lines were available (all Finnish patients) did not show a size difference of the transcript or altered level of expression when compared to control subjects. All the mutations cosegregated with the disease in the respective families and were predicted to result in truncation of the conceptual protein (Table 1). This provides strong evidence that alterations of the APGD1 gene represent the primary cause for the APECED disease.

Please replace the paragraph starting on page 27, line 31, with the following amended paragraph:

Polyclonal antibodies against the AIRE protein were obtained by injecting rabbits with the synthetic peptides MATDAALRRLLRLHR (SEQ ID NO:19) (corresponding to aa 1-15) and SQPRKGRKPPAVPK (SEQ ID NO:20) (corresponding to aa 107-120), respectively. The resulting immune sera sp97179 (for aa 1-15) and sp97181 (for aa 107-120) were affinity purified against their corresponding synthetic peptides immobilized on a HiTrap NHS-activated 1 ml column (Pharmacia) according to the manufacturer's recommendations.

Please replace the paragraph starting on page 35, line 12, with the following amended paragraph:

Human: RT-PCR analysis was performed on Clontech's Human Immune System Multiple Tissue cDNA Panel of first-strand cDNA from the following tissues: human bone marrow, fetal liver, lymph node, peripheral blood leukocyte, spleen, thymus and tonsil. Primers B127FR4-21 (5'-GGC TTC TGA GGC TGC ACC) (SEQ ID NO:28) and B127FR4-29 (5'-GCT CTG GAT GGC CTA CTG C) (SEQ ID NO:21) were used to amplify a 1.6 kb region specific for h*AIRE*. Each PCR was performed in a 50 ml reaction mix containing 5 ml of MTC Panel cDNA, 10-20 pmol of each primer, 1 ml of a 10 mM dNTP mix, 5 ml of Perkin Elmer *GeneAmp*" 10X-PCR buffer (100 mM Tris-HCl pH 8.3; 500 mM KCl; 15 mM $MgCl_2$; 0.01% w/v gelatin), and 3 ml of freshly prepared 28:1 (7 mM:1.4 mM) mixture of TaqStart Antibody (Clontech) and *AmpliTaq*" DNA Polymerase (Perkin Elmer). PCR reactions were performed in a Biometra UNO II thermocycler beginning with a 2 min initial denaturation step at 94°C, followed by 38 cycles of 94°C for 45 sec, 56°C for 40 sec, 72°C for 1 min, and a final extension step at 72°C for 5 min. Products of the PCR were re-amplified with nested primers B127FR4-17 (5'-AGA AGT GCA TCC AGG TTG GC) (SEQ ID NO:22) and B127FR4-33 (5'-GTG TGC TCG CTC AGA AGG G) (SEQ ID NO:23) to confirm that the products were specific to h*AIRE*.

Please replace the paragraph starting on page 35, line 30, with the following amended paragraph:

Mouse: Mouse primers Mforw4 (5'-TGG CAG GTG GGG ATG GAA) (SEQ ID NO:24) and Mrev15 (5'-GGA GGG ATG GAA GGG GAG GA) (SEQ ID NO:25) were used to amplify AIRE specific regions from Clontech's Mouse Multiple Tissue cDNA Panel 1 (consisting of first-strand cDNA from mouse heart, brain, spleen, lung, liver, skeletal, kidney, testis and 7-day, 11-day, 15-day and 17-day embryo tissues). PCR reaction mixtures were set

up according to the same conditions described for human RT-PCR's, with the exception of using mouse specific primers and a PCR annealing temperature of 63°C.

Please replace the paragraph starting on page 36, line 8, with the following amended paragraph:

Chromosomal localization of m*AIRE* was established by PCR analysis of mouse chromosomes 3, 10 and 17. PCR amplifications were performed using mouse specific primers Mforw2 (5'-TCC CAC CTG AAG ACT AAG C) (SEQ ID NO:26) and Mrev32 (5'-TCA CAG CTC TCT GGA CAG AA) (SEQ ID. NO.: 27) on cell hybrids SN11CS3 (chromosome 3), SN17C3 (chromosome 10) and EJ167 (chromosomes 17 and 3 on a human background). PCR reactions were performed in 30 ml volumes containing 5 ml of mouse chromosomal preparations, 10-20 pmol of each primer, 1 ml of a 10 mM dNTP mix, 5 ml of Perkin Elmer *GeneAmp*" 10X-PCR buffer, and 3 ml of freshly prepared 28:1 (7 mM:1.4 mM) mixture of TaqStart Antibody (Clontech) and *AmpliTaq*" DNA Polymerase (Perkin Elmer). PCR reactions were performed in a Biometra UNO II thermocycler beginning with a 2 min initial denaturation step at 94°C, followed by 35 cycles of 94°C for 45 sec, 51°C for 40 sec, 72°C for 2 min, and a final extension step at 72°C for 5 min.

Please replace the paragraph starting on page 37, line 4, with the following amended paragraph:

Computer Analysis: Genome-wide repeats were identified with the Repeatmasker program (A.F.A Smit and P. Green on the internet in file transfer protocol (ftp) at ~~http://ftp.genome.washington.edu/RM/RepeatMasker.html~~ genome.washington.edu/RM/RepeatMasker.html). The GC content and distribution was determined with the LPC algorithm (Huang 1994). Homology searches against various databases were performed using BLAST version 1.4 (Altschul et al. 1990) and FASTA version 2.0 (Pearson and Lipman 1998). Programs GRAIL2 (Uberbacher and Mural 1991), XPOUND (Thomas and Skolnick 1994), MZEF (Zhang 1997) and GENSCAN (Burge and Karlin 1997)

were used for exon prediction. Promoter predictions were done with „Promoter Scan II" (Prestridge 1995) and „Transcription Start Site" using both Ghosh/Prestridge (TSSG) and Wigender (TSSW) motif databases (V.V. Solovyev, A.A. Salamov and C.B. Lawrence on the internet at ~~http://dot.imgen.bcm.tmc.edu:9331/gene-finder/gf.html~~ dot.imgen.bcm.tmc.edu:9331/gene-finder/gf.html).

Please replace the paragraph starting on page 37, line 15, with the following amended paragraph:

Cosmid L12287 was completely sequenced (~~46,8872~~ 468,872 bp long; EMBL accession no. AF073797) and the data were compared with the human *AIRE* gene locus that we have previously sequenced (36,284 bp, accession no. HSAJ9610). Automatic sequence analysis of clone L12287 was performed with the Rummage software (~~http://www.genome.imb_jena.de~~ available on the worldwide web at genome.imb_jena.de). Gene prediction programs detected the *AIRE* gene and revealed also an incomplete gene model located 6 kb from the 5' end of *AIRE* that was corroborated by anonymous EST matches (e.g. accession no. AA413561). Interestingly, one of the anonymous exons showed high homology with a trapped exon (HC21EXc32; D86111) mapping to human chromosome 21q22.3 (Genebank Accession no. D86111) This confirmed the high degree of conserved syntheny between mouse and human in this region.